Active Inference

Melih Kandemir

University of Southern Denmark
Department of Mathematics and Computer Science (IMADA)
mymail@imada.sdu.dk

October 7, 2024

The brain and Bayes

- Brain as a predictive machine (Kant's second Copernican revolution)
- Perception as unconscious inference (Helmholtz, 1866)
- Bayesian brain hypothesis (Doya, 2007)
  - Perception is not a purely bottom-up transduction of sensory states into internal representations
  - It is an inferential process that combines
    - (Top-down) Prior info about most likely causes of sensations
    - (Bottom-up) Sensory stimuli
- Perception is a constructive inside-out process
  - The brain uses sensations to confirm/disconfirm hypotheses about how they were generated

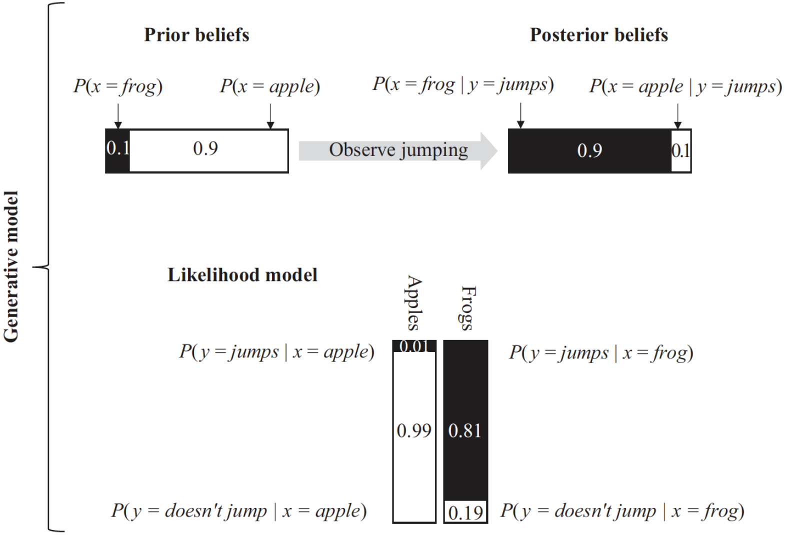
Translating Bayesian terms to the brain language

Forward model:

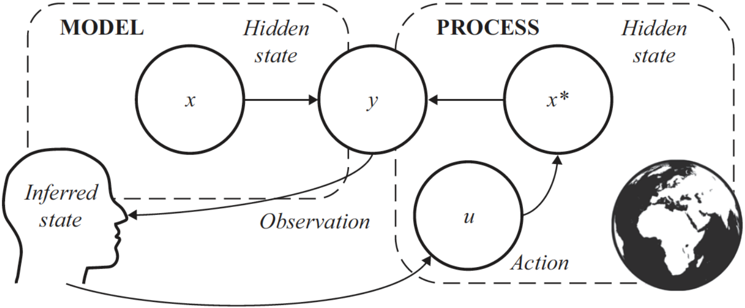
- $P(x)$: Organism's guess about a concept, i.e. a perceptual regularity
- $P(y|x)$: Organism's knowledge about how concepts relate to sensory inputs

Surprise: $-\log P(y) \Rightarrow$ organism cannot explain environment $\Rightarrow$ danger $\Rightarrow$ model update
Bayesian surprise: How much belief update required given an observation
$D_{KL}[P(x|y)||P(x)] = E_{x \sim P(x|y)}[\ln P(x|y) - \ln P(x)]$.

The Bayesian brain example

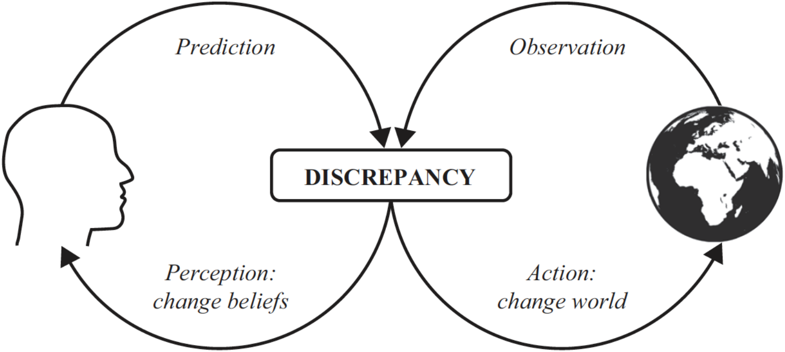


Perception as inference



Generative model $\neq$ Generative process
Generative model is subjective and limited by the computational and mnemonic resources of the organism.

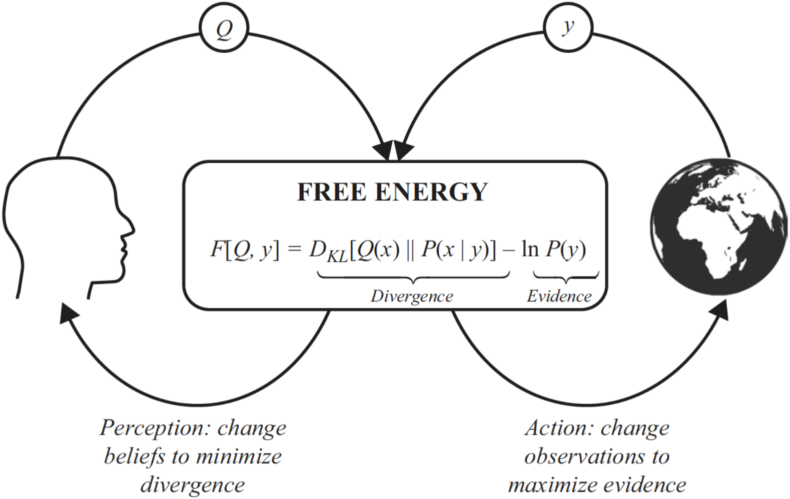
Actions for optimal belief updates



What does the ELBO mean to a brain?

$$
\begin{aligned}
F[Q,y] &= \underbrace{-\mathbb{E}_{Q(x)}[\ln P(y,x)]}_{\textit{Energy}} - \underbrace{H[Q(x)]}_{\textit{Entropy}} \\
&= \underbrace{D_{KL}[Q(x) \,||\, P(x)]}_{\textit{Complexity}} - \underbrace{\mathbb{E}_{Q(x)}[\ln P(y|x)]}_{\textit{Accuracy}} \\
&= \underbrace{D_{KL}[Q(x) \,||\, P(x|y)]}_{\textit{Divergence}} - \underbrace{\ln P(y)}_{\textit{Evidence}}
\end{aligned}
$$

How does the ELBO relate to the outside world?



Expected Free Energy: When actions come to play

$$
\begin{aligned}
G(\pi) &= -\underbrace{\mathbb{E}_{Q(\tilde{x},\tilde{y}|\pi)}[D_{KL}[Q(\tilde{x}|\tilde{y},\pi)\,||\,Q(\tilde{x}|\pi)]]}_{\text{Information gain}} - \underbrace{\mathbb{E}_{Q(\tilde{y}|\pi)}[\ln P(\tilde{y}|C)]}_{\text{Pragmatic value}} \\
&= \underbrace{\mathbb{E}_{Q(\tilde{x}|\pi)}[H[P(\tilde{y}|\tilde{x})]]}_{\text{Expected ambiguity}} + \underbrace{D_{KL}[Q(\tilde{y}|\pi)\,||\,P(\tilde{y}|C)]}_{\text{Risk (outcomes)}} \\
&\leq \underbrace{\mathbb{E}_{Q(\tilde{x}|\pi)}[H[P(\tilde{y}|\tilde{x})]]}_{\text{Expected ambiguity}} + \underbrace{D_{KL}[Q(\tilde{x}|\pi)\,||\,P(\tilde{x}|C)]}_{\text{Risk (states)}} \\
&= \underbrace{-\mathbb{E}_{Q(\tilde{x},\tilde{y}|\pi)}[\ln P(\tilde{y},\tilde{x}|C)]}_{\text{Expected energy}} - \underbrace{H[Q(\tilde{x}|\pi)]}_{\text{Entropy}}
\end{aligned}
$$

$$
Q(\tilde{x},\tilde{y}|\pi) \triangleq Q(\tilde{x}|\pi)P(\tilde{y}|\tilde{x})
$$

What does Expected Free Energy mean to a brain?

EXPECTED FREE ENERGY

$$G(\pi) = -\mathbb{E}_{Q(\tilde{y}\mid\pi)}\big[\underbrace{\ln P(\tilde{y}\mid C)}_{1}\big] - D_{KL}\big[\underbrace{Q(\tilde{y}\mid\tilde{x})}_{2}\,\underbrace{Q(\tilde{x}\mid\pi)}_{3} \,\|\, \underbrace{Q(\tilde{y}\mid\pi)}_{4}\,\underbrace{Q(\tilde{x}\mid\pi)}_{5}\big]$$

2 3 4 5

$$D_{KL}\big[Q(\tilde{y}\mid\tilde{x})\,Q(\tilde{x}\mid\pi) \,\|\, Q(\tilde{y}\mid\pi)\,Q(\tilde{x}\mid\pi)\big]$$

$$= \mathbb{E}_{Q(\tilde{y}\mid\pi)}\big[D_{KL}\big[Q(\tilde{x}\mid\tilde{y},\pi) \,\|\, Q(\tilde{x}\mid\pi)\big]\big]$$

BAYESIAN SURPRISE
OPTIMAL BAYESIAN DESIGN
INTRINSIC MOTIVATION
INFOMAX PRINCIPLE

1 3 4 5

$$D_{KL}\big[Q(\tilde{y}\mid\pi) \,\|\, P(\tilde{y}\mid C)\big]$$

RISK-SENSITIVE POLICIES
KL CONTROL

1

$$\mathbb{E}_{Q(\tilde{y}\mid\pi)}\big[\ln P(\tilde{y}\mid C)\big]$$

BAYESIAN DECISION THEORY
EXPECTED UTILITY THEORY